HAL open science

# Breaking the Bounds: Introducing Informed Spectral Analysis

Sylvain Marchand, Dominique Fourer

## To cite this version:

Sylvain Marchand, Dominique Fourer. Breaking the Bounds: Introducing Informed Spectral Analysis. International Conference on Digital Audio Effects (DAFx), Sep 2010, Graz, Austria. pp.359-366. hal-00523321

**HAL Id: hal-00523321**

**https://hal.science/hal-00523321v1**

Submitted on 4 Oct 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# BREAKING THE BOUNDS: INTRODUCING INFORMED SPECTRAL ANALYSIS

*Sylvain Marchand*
LaBRI – CNRS
University of Bordeaux 1
Talence, France
sylvain.marchand@labri.fr

*Dominique Fourer*
LaBRI – CNRS
University of Bordeaux 1
Talence, France
dominique.fourer@labri.fr

**ABSTRACT**

Sound applications based on sinusoidal modeling highly depend on the efficiency and the precision of the estimators of its analysis stage. In a previous work, theoretical bounds for the best achievable precision were shown and these bounds are reached by efficient estimators like the reassignment or the derivative methods. We show that it is possible to break these theoretical bounds with just a few additional bits of information of the original content, introducing the concept of "informed analysis". This paper shows that existing estimators combined with some additional information can reach any expected level of precision, even in very low signal-to-noise ratio conditions, thus enabling high-quality sound effects, without the typical but unwanted musical noise.

## 1. INTRODUCTION

For decades, researchers have spent lots of efforts improving the precision of sound analysis, since it is often crucial for the perceived quality of the sound transformations. And yet this quality is not sufficient for demanding applications.

One approach is to try to improve the analysis methods even further, without guarantee of success though. Indeed, theoretical bounds may exist, indicating the minimal error (*i.e.* maximal quality) reachable without extra information (blind approach).

Another approach is to inject some information. This can be prior knowledge about the sound sources and / or the way the human auditory system will perceive them (computational auditory scene analysis approach). But this information can also come from a manual annotation (semi-automatic analysis). Moreover, when access to the compositional process is given, another option is to use some bits of the ground truth as additional information in order to help the analysis process. This is the concept of "informed analysis" (in opposition to the blind approach), used recently to improve sound source separation [1]. Of course, using the whole information would be cheating. Here, we intend to determine and use only the minimal amount of information.

Where this information should be added then? It could be anywhere in the analysis / transformation / synthesis chain. We would like to quantify the minimal amount of information necessary to achieve a specific task (*e.g.* some audio effect) with a desired level of quality. The measurement of the quality could be done at the end of the processing chain, using either objective measurements based on signal-to-noise ratios or subjective listening tests.

Here we want to measure and improve the quality of the sinusoidal modeling approach, of great interest for sound transformations as shown by McAulay and Quatieri [2] for speech signals and Serra and Smith [3] for musical sounds. We start by the analysis stage, and we want to evaluate its quality objectively.

Sinusoidal analysis usually consists of two steps: peak picking (short-term analysis) and partial tracking (long-term analysis). Whereas evaluating partial tracking is still an open issue [4], the evaluation of the peak picking has a clear theoretical background, since the Cramér-Rao lower bounds (CRBs) give the minimal error variance for unbiased estimators of the peak parameters.

To go below these bounds, the only solution is to add some information. An interesting approach could be to use prior knowledge about the sound structure, giving constraints among the parameters of the different sinusoids. Here, we prefer to stay at the most general level, with no priors on the sound sources. We will take advantage of a few bits of the ground truth to improve the quality of the estimation of the parameters of each sinusoid.

The present work should be regarded as a proof of concept of informed sound analysis. We focus on the estimation of the frequency of the sinusoids, a well-studied problem. Even with a rather naive approach we show that we can go below the theoretical bounds, which means that we are able to extract the sinusoids with a much better precision than without information, allowing musical applications with a sufficient quality.

Starting with the most studied problem (frequency estimation) of the first analysis step (peak picking) of the sinusoidal modeling chain, we expect that the precision gain should propagate through the chain and improve the overall quality. For example, accurate spectral peaks would ease the partial tracking step, and so on.

Of course, as a long-term research, we should optimize qualitatively and quantitatively the additional information at each stage of the complete analysis / transformation / synthesis chain.

This paper is organized as follows. Section 2 explains the classic frequency estimation using the reassignment method and recalls the theoretical lower bounds for the estimation error. Section 3 then presents our new approach, namely "informed estimation", consisting in taking advantage of a few bits of additional information to derive a more precise estimate. This improvement is illustrated by our experimental results in Section 4.

## 2. CLASSIC FREQUENCY ESTIMATION

Sinusoidal modeling involves a complete analysis / transformation / synthesis chain. The estimation of the frequency of each sinusoid is often the very first step. Thus, for high-quality sound transformations, the precision of this estimation is crucial in practice. Yet, theoretical considerations show that this precision is limited.

### 2.1. Sinusoidal Model

Additive synthesis can be considered as a spectrum modeling technique. It is originally rooted in Fourier's theorem, which states

that any periodic function can be modeled as a sum of sinusoids at various amplitudes and harmonically related frequencies. In this paper we consider the sinusoidal model under its most general expression, which is a sum of complex exponentials (the *partials*) with time-varying amplitudes $a_p$ and non-harmonically related frequencies $\omega_p$ (defined as the first derivative of the phases $\phi_p$). The resulting signal $s$ is thus given by:

$$s(t) = \sum_{p=1}^{P} a_p(t) \exp(j\phi_p(t)). \tag{1}$$

The amplitudes and frequencies may evolve within an analysis frame (non-stationary case) under first-order amplitude and frequency modulations. Furthermore, as the present study focuses on frequency precision (*i.e.* accurately measuring this parameter) rather than frequency resolution (*i.e.* resolving closely-spaced sinusoids), the signal model is reduced to only one partial ($P = 1$). The subscript notation for the partials is then useless. We define then $\Pi_0$ as being the value of the parameter $\Pi$ at time 0, corresponding to the center of the analysis frame. The signal $s$ is then given by:

$$s(t) = \exp\left( \underbrace{(\lambda_0 + \mu_0 t)}_{\lambda(t)=\log(a(t))} + j \underbrace{\left(\phi_0 + \omega_0 t + \frac{\psi_0}{2} t^2\right)}_{\phi(t)} \right) \tag{2}$$

where $\mu_0$ (the amplitude modulation) is the derivative of $\lambda$ (the log-amplitude), and $\omega_0$ (the frequency), $\psi_0$ (the frequency modulation) are respectively, the first and second derivatives of $\phi$ (the phase). Thus, the log-amplitude and the phase are modeled by polynomials of degrees 1 and 2, respectively (see [5] for the corresponding synthesis method). These polynomial models can be viewed either as truncated Taylor expansions of more complicated amplitude and frequency modulations (*e.g.* tremolo / vibrato), or either as an extension of the stationary case (where $\mu_0 = \psi_0 = 0$).

### 2.2. Reassignment Method

This paper focuses on the estimation of the frequency $\omega_0$. As mentioned in [6, 7], there are many estimators. Many efficient ones are based on the short-time Fourier transform (STFT):

$$S_w(t, \omega) = \int_{-\infty}^{+\infty} s(\tau) w(\tau - t) \exp\left(-j\omega(\tau - t)\right) \, d\tau \tag{3}$$

where $S_w$ is the short-time spectrum of the signal $s$. This involves an analysis window $w$, band-limited in such a way that for any frequency corresponding to one specific partial (corresponding to some local maximum in the magnitude spectrum), the influence of the other partials can be neglected (in the general case when $P > 1$). We use the zero-centered (symmetric) Hann window of duration $T$, defined on the $[-T/2; +T/2]$ interval by:

$$w(t) = \frac{1}{2}\left(1 + \cos\left(2\pi \frac{t}{T}\right)\right). \tag{4}$$

The reassignment method, first proposed by Kodera, Gendrin, and de Villedary [8, 9], was generalized by Auger and Flandrin [10] for time and frequency. By considering Equation (3), one can easily derive:

$$\frac{\partial}{\partial t} \log\left(S_w(t, \omega)\right) = j\omega - \frac{S_{w'}(t, \omega)}{S_w(t, \omega)} \tag{5}$$

where $w'$ denotes the derivative of $w$. Since the frequency is the derivative of the phase $\phi = \Im(\log(S_w))$, we then obtain the reassigned frequency $\hat{\omega}$:

$$\hat{\omega}(t, \omega) = \frac{\partial}{\partial t} \Im\left(\log\left(S_w(t, \omega)\right)\right) = \omega - \Im\left(\frac{S_{w'}(t, \omega)}{S_w(t, \omega)}\right). \tag{6}$$

In practice, for a partial $p$ corresponding to a local maximum $m$ of the (discrete) magnitude spectrum at the (discrete) frequency $\omega_m$, the estimate of the frequency is given by $\hat{\omega}_0 = \hat{\omega}(0, \omega_m)$.

The reassignment method seems currently the best STFT-based method in terms of efficiency and estimation precision, at least regarding frequency (see [6]). The generalized derivative method [7] could be used too, as well as high-resolution methods [11]. In all cases, the precision is close to optimal. The high-resolution methods improve the frequency resolution, but not the estimation precision, which is always limited by the Cramér-Rao lower bound.

### 2.3. Theoretical Bound

In practice, we consider discrete-time signals $s$, with sampling rate $F_s$, consisting of 1 complex exponential generated according to Equation (2) with an initial amplitude $a_0$, and mixed with a Gaussian white noise of variance $\sigma^2$ – the signal-to-noise ratio (SNR) is then $a_0/\sigma$. To make the parameters independent of the sampling frequency, in the remainder of this paper we normalize $\mu$ and $\omega$ (by $F_s$).

The analysis frames we consider are of odd length $N = 2H + 1$ samples (the duration of the analysis window being $T = N/F_s$), with the estimation time 0 set at their center. In Equation (3), the continuous integral turns into a discrete summation over $N$ values, with an index from $-H$ to $+H$ (the fast Fourier transform is used).

When evaluating the performance of an estimator in the presence of noise and in terms of the variance of the estimation error, an interesting element to compare with is the Cramér-Rao bound (CRB). The CRB is defined as the limit to the best possible performance achievable by an unbiased estimator given a data set. For the model of Equation (2), for the five model parameters, these bounds have been derived by Zhou *et al.* [12]. We will restrict our study to the frequency parameter, and consider the asymptotic version (for a large $N$ and a high number of observations) of the corresponding bound.

Djurić and Kay [13] have shown that the CRB depends on the time sample $n_0$ at which the parameters are estimated, and that the optimal choice in terms of lower bounds is to set $n_0$ at the center of the frame, *i.e.* $n_0 = H$, since the CRB depends on:

$$\epsilon_k(\mu, N) = \sum_{n=0}^{N-1} \left(\frac{n - n_0}{N}\right)^k \exp\left(2\mu \frac{n - n_0}{N}\right). \tag{7}$$

As explained by Zhou *et al.* [12], the expression of the bound is different whether there is a frequency modulation or not (because this changes the degree of the polynomial associated to the phase). In the absence of frequency modulation ($\psi = 0$), the lower bound for the frequency $\omega$ is given by:

$$\text{CRB}(N, \sigma, a_0, \mu) \approx \frac{\sigma^2 \epsilon_0}{2a_0^2 N^2 (\epsilon_0 \epsilon_2 - {\epsilon_1}^2)}, \tag{8}$$

whereas in the presence of frequency modulation, it turns into:

$$\text{CRB} \approx \frac{\sigma^2 (\epsilon_0 \epsilon_4 - {\epsilon_2}^2)}{2a_0^2 N^2 (\epsilon_0 \epsilon_2 \epsilon_4 - {\epsilon_1}^2 \epsilon_4 - \epsilon_0 {\epsilon_3}^2 + 2\epsilon_1 \epsilon_2 \epsilon_3 - {\epsilon_2}^3)}. \tag{9}$$

Thus, the precision of the estimation of the frequency of each sinusoid is limited by this CRB, at least without using additional information...

## 3. INFORMED FREQUENCY ESTIMATION

Now, $\omega$ is a frequency parameter and $\hat{\omega}$ is its estimation – done using the classic estimation method (see Section 2). We are aiming at reducing the estimation error $|\omega - \hat{\omega}|$ by introducing a few bits of the representation of the exact value $\Omega$ at the best place in the representation of the estimated value $\hat{\Omega}$. This will lead to a frequency estimation method which is informed by the additional bits, and thus more precise.

### 3.1. Coding Convention

Let $\omega$ be a normalized frequency, thus within the $[0; .5)$ interval. Let us consider then its $k$-length fixed-point binary representation, denoted by $\Omega = (\Omega_1, \Omega_2, \cdots, \Omega_k)$, $\Omega(i)$ denoting $\Omega_i$, obtained with the standard binary coding application $\mathcal{C} : \mathbb{R} \rightarrow \{0,1\}^k$, thus:

$$\Omega(i) = \lfloor \omega \cdot 2^{(i+1)} \rfloor \bmod 2 \tag{10}$$

where $\lfloor x \rfloor$ denotes the integral part of $x$. Each coded value $\Omega$ can thus be decoded with the standard binary decoding function:

$$\mathcal{D}(\Omega) = \sum_{i=1}^{k} \Omega(i) \cdot 2^{-(i+1)} \tag{11}$$

and, in these conditions, we have $|\omega - \mathcal{D}(\mathcal{C}(\omega))| < 2^{-(k+1)}$ (*i.e.* the limit of the precision due to the quantification using the $k$ bits).

### 3.2. Information Extraction

For a given $\omega$, the most informative area starts at the most significant bit of the error $|\omega - \hat{\omega}|$. Indeed, replacing this first erroneous bit could divide the error by 2. Then, replacing in turn each subsequent bit could produce a comparable enhancement. However, this approach is not realistic since it would make the error correcting algorithm dependent on the frequency $\omega$ (which is unknown).

However, for a given noise variance $\sigma^2$, we can study the distribution of the most significant bit of the error $|\omega - \hat{\omega}|$ (for all possible $\omega$). We define $I_\sigma$ the index of the most significant bit (and with a significant number of occurrences). As shown in Figures 1 and 2, this index is close to the one of the mode of the distribution, and $I_\sigma$ is a growing function of the noise parameter $\sigma$. Once $I_\sigma$ is known, the procedure for extracting $\mathcal{I}$ containing the $d$ bits of information is almost straightforward:

$C \leftarrow \mathcal{C}(\omega)$
$\mathcal{I}(1:d) \leftarrow C(I_\sigma - 1 : I_\sigma + d - 2)$

except that we choose to extract the $d$-bit sub-code from 1 bit before $I_\sigma$, to be able to handle special cases (see below).

### 3.3. Error Correction

In practice, $\sigma$ is unknown and has to be estimated. We could use the method proposed in [14], since this method tends to overestimate the noise level which will result in an underestimation of $I_\sigma$. This will lead to a loss of efficiency of the information, but the results will be at least as good as with the classic estimation.

Overestimating $I_\sigma$ would be dangerous. Indeed, the first $I_\sigma - 1$ bits must be reliable, otherwise the informing bits might lead to a wrong representation possibly producing an important error (thus decreasing the estimation precision instead of increasing it).

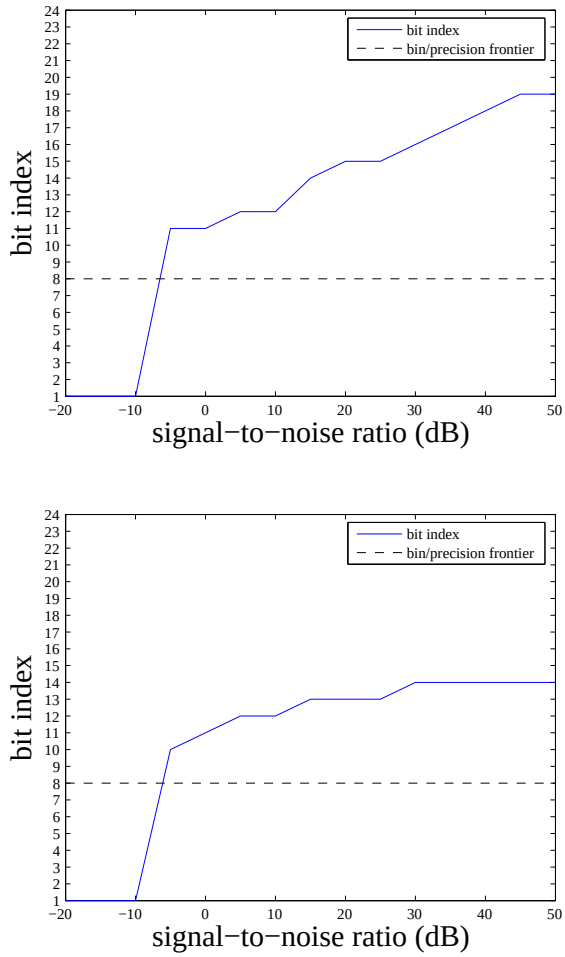



Figure 2: *Index $I_\sigma$ of the most significant bit of the estimation error $|\omega - \hat{\omega}|$ of the reassignment method (window size $N = 511$) as a function of the signal-to-noise ratio, in the stationary (top) and non-stationary (bottom) cases.*

In the present work, $\sigma$ is supposed to be known. The design of a complete method including the estimation of $\sigma$ is part of our future research.

We define the $d$-informed coded estimation of $\omega$ from the representation of $\hat{\omega}$ as follows:

$$\tilde{\Omega} = \left( \underbrace{\hat{\Omega}_1, \cdots, \hat{\Omega}_{I_\sigma - 1}}_{\hat{\Omega}(1:I_\sigma - 1)}, \underbrace{\mathcal{I}(2), \cdots, \mathcal{I}(d)}_{\mathcal{I}(2:d)}, \underbrace{\hat{\Omega}_{I_\sigma + d - 1}, \cdots, \hat{\Omega}_k}_{\hat{\Omega}(I_\sigma + d - 1:k)} \right) \tag{12}$$

and the informed estimation of the frequency is $\tilde{\omega} = \mathcal{D}(\tilde{\Omega})$.

Whereas the informing bits are reliable, we might deal with the case of $\exists i \in (0; I_\sigma), \hat{\Omega}(i) \neq \Omega(i)$. In this case, modifying the bit values placed after $I_\sigma$ may increase the error $|\omega - \tilde{\omega}|$, above the original estimation error $|\omega - \hat{\omega}|$. Indeed, two close values can have very different representations. This is the case for

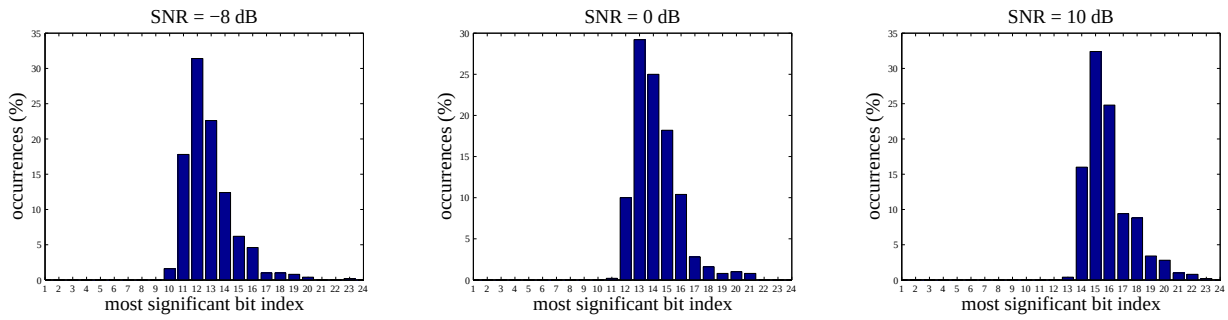



Figure 1: *Histograms of the most significant bit index of the estimation error $|\omega - \hat{\omega}|$ of the reassignment method (window size $N = 511$) for several signal-to-noise ratios (SNR = -8dB, 0dB, and 10dB) in the stationary case and with a precision of $k = 24$ bits.*

$X = (0, 0, 1, 1, 1)$ and $Y = (0, 1, 0, 0, 0)$. The binary representations are almost completely different, although we obtain $Y$ by incrementing $X$ only by 1. Fortunately, by definition of $I_\sigma$, we have:

$$|\omega - \hat{\omega}| < 2^{-I_\sigma} \tag{13}$$

and thus the $I_\sigma - 1$ first bits of the representation of $\omega$ and $\hat{\omega}$ can only differ by 1. This happens if and only if $\Omega(I_\sigma - 1) \neq \hat{\Omega}(I_\sigma - 1)$, that is $\mathcal{I}(1) \neq \hat{\Omega}(I_\sigma - 1)$.

Finally, the algorithm taking advantage of the information $\mathcal{I}$ to compute a more precise estimate of the frequency $\tilde{\omega}$ is:

$\Omega \leftarrow \mathcal{C}(\hat{\omega})$
$\Omega(I_\sigma : I_\sigma + d - 2) \leftarrow \mathcal{I}(2 : d)$
$\tilde{\omega} \leftarrow \mathcal{D}(\Omega)$
**if** $\mathcal{I}(1) \neq \Omega(I_\sigma - 1)$ **then**
  $\Omega^{\text{ante}} \leftarrow \Omega(1 : I_\sigma - 1)$
  $\Omega^{\text{post}} \leftarrow \Omega(I_\sigma : k)$
  $\omega^+ \leftarrow \mathcal{D}(\text{inc}(\Omega^{\text{ante}}), \Omega^{\text{post}})$
  $\omega^- \leftarrow \mathcal{D}(\text{dec}(\Omega^{\text{ante}}), \Omega^{\text{post}})$
  **if** $|\hat{\omega} - \omega^+| < |\hat{\omega} - \omega^-|$ **then**
    $\tilde{\omega} \leftarrow \omega^+$
  **else**
    $\tilde{\omega} \leftarrow \omega^-$
  **end if**
**end if**

where inc and dec stands for incrementing and decrementing the binary representation, respectively.

For more robustness, we have also to consider the eventuality when $I_\sigma$ is overestimated, resulting in the violation of the Equation (13) by both $\omega^+$ and $\omega^-$. This problematic case is easily detected when $|\tilde{\omega} - \hat{\omega}| \geq 2^{-I_\sigma}$. Then, it is safer to revert to the $\hat{\omega}$ estimate.

Thanks to this algorithm, the first $I_\sigma + d - 2$ bits of $\tilde{\Omega}$ are now reliable, *i.e.* $\tilde{\Omega}(i) = \Omega(i)$ for $i < I_\sigma + d - 1$. Thus $\tilde{\omega}$ gets closer to $\omega$ as $d$ grows, and the estimation precision increases.

### 3.4. Theoretical Bound

We can also define a theoretical lower bound in the informed case, supposing the suitability of each additional bit of information. Indeed, in the best case, each bit of information is able to divide the standard deviation of the error by 2. Thus, for a number of $d$ informing bits, an informed lower bound (ILB) for the variance of the error is given by:

$$\text{ILB}(d, N, \sigma, a, \mu) = \text{CRB}(N, \sigma, a, \mu)/2^{2 \cdot d}. \tag{14}$$

In fact, the ILB is limited by the $k$-bit precision of the binary representation, unless the CRB itself falls below this precision. This gives an ILB consisting of 3 line segments, as shown in Figure 5:

1. the first one given by Equation (14), where the $d$ bits of information are not enough to reach the $k$-bit precision limit;
2. the second (horizontal) segment where less than $d$ bits are needed to reach the (constant) $k$-bits precision limit;
3. the third segment where the CRB allows to exceed the $k$-bit precision without any information (the estimation precision being then better than the representation precision).

## 4. PRACTICAL EXPERIMENTS AND RESULTS

With the informed approach, we can enhance the estimation precision even below the CRB (down to the ILB). We demonstrate this with numerical simulations. In practice, we are now able to reach any level of precision, at the expense of additional informing bits. The frequency trajectories of the partials are then much more accurate, and the unwanted musical noise due to the estimation errors is now inaudible.

### 4.1. Simulation Results

Let us first consider a discrete-time signal $s$ with sampling rate $F_s = 44100$Hz, consisting of 1 complex exponential of amplitude $a_0 = 1$ according to Equation (2) plus a Gaussian white noise of variance $\sigma^2$. The signal-to-noise ratio (SNR) expressed in dB is $10\log_{10}(a_0^2/\sigma^2)$ and goes from $-20$dB to $+50$dB by steps of 5dB. We consider frames of size $N = 511$, and we use the Hann analysis window $w$, defined for continuous time by Equation (4).

For each SNR, we consider 99 frequencies ($\omega_0$) linearly distributed in the $(0, F_s/2)$ interval, and 9 phases ($\phi_0$) linearly distributed in $(-\pi, +\pi)$. The amplitude modulation ($\mu_0$) is either 0 (stationary case) or one of 5 values linearly distributed in the $[-100, +100]$ interval (non-stationary case). The frequency modulation ($\psi_0$) is either 0 (stationary case) or one of 5 values linearly distributed in $[-10000, +10000]$ (non-stationary case).

The values for $I_\sigma$ (see Figure 2) were statistically estimated from 500 random values uniformly distributed in the ranges above.

We then compare the classic reassignment method (Section 2) with its informed variant (Section 3), and plot the Cramér-Rao and informed lower bounds. We consider two situations: either a fixed number $d = 5$ of informing bits (Figure 5) or all the necessary bits

to reach the target precision corresponding to $k = 16$ or 24 bits (Figure 6).

When looking at the results of these experiments, as expected the informed reassignment exhibits a variance of the estimation error lower than CRB in every case.

Below $-10$dB (high-error range), the noise conditions are so bad that $I_\sigma = 1$ and every bit of information is useful.

Above $-10$dB, our informed method is less efficient because from time to time some informing bits are useless. So it is more difficult to reach the ILB (see Figure 5), although this is possible if the whole information can be used (see Figure 6).

It is also interesting to notice the robustness of the informing algorithm, which also works in the non-stationary case.

### 4.2. Vibrato Sound Experiment

The second experiment consists in estimating the (non-stationary) frequency trajectory of a partial with a vibrato as described in Figure 3. We clearly see that the trajectory obtained with the classic reassignment method is erroneous. These errors may result in an annoying musical noise. However, when informing the estimation up to 16 bits, the error is reduced and this noise becomes inaudible.

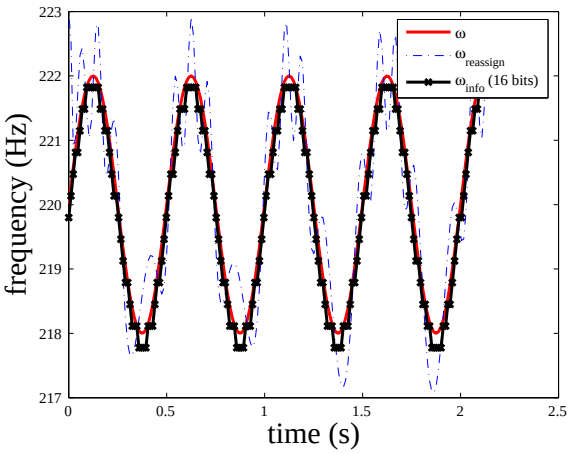



Figure 3: *Frequency estimation of a non-stationary sinusoid of frequency* $\omega(t) = 220+2\sin(2\pi\cdot 2t)$ *(in Hz), using the classic and informed versions of the reassignment method. The classic method obtains an irregular trajectory due to estimation error whereas the informed method trajectory is more reliable but quantified.*

### 4.3. Natural Sound Experiment

The third experiment considers a 220-Hz piano tone of approximately 2 seconds. We first obtained the reference sinusoidal parameters from a real piano tone by the classic estimation in the absence of noise (so that the estimation is almost perfect). Second, the reference tone is obtained from these parameters by classic additive synthesis (with linear amplitude and phase interpolations).

Then a Gaussian noise with fixed variance $\sigma^2$ is added to the reference tone. Since we know $\sigma$ and the amplitude of each partial, the signal-to-noise conditions are known (see Figure 4), so that we can extract all the needed information $\mathcal{I}$ to reach the desired precision ($k = 16$ bits) with our new method.

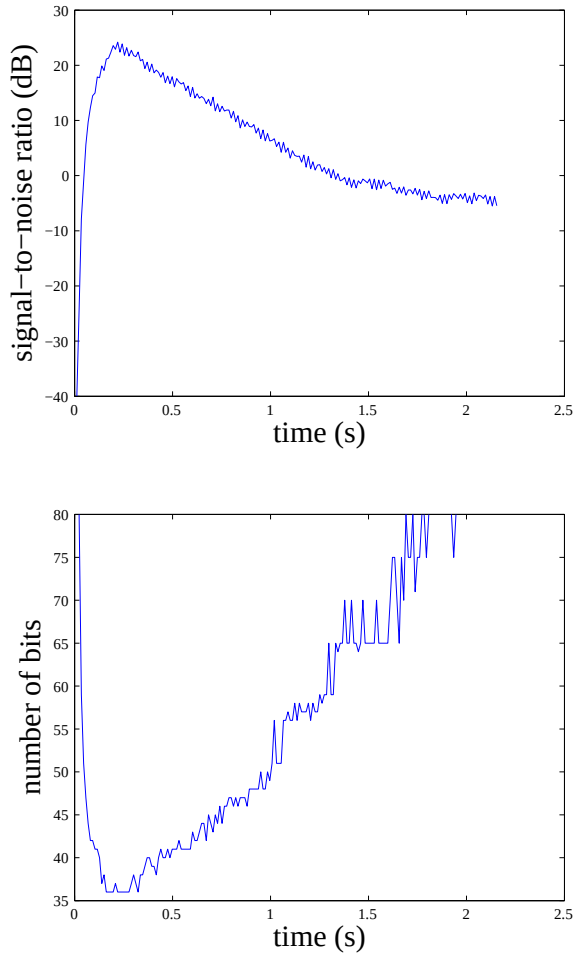



Figure 4: *Overall signal-to-noise ratio (top) and instantaneous bit rate (bottom) required to inform a sound of piano (with* 5 *partials).*

We use then the classic and informed versions of the reassignment method to estimate the sinusoidal model parameters, and resynthesize the sounds using the classic additive synthesis.

As shown by informal listening tests[1], whereas an annoying musical noise can be heard with the sound obtained by the classic method, the one obtained with the informed method is similar to the reference sound.

This piano tone consists of 5 partials during 187 frames (approx. 2.16s), and thus represents $187\times N\times 16 = 1528912$ bits at CD quality. The total information for the partials frequency at a precision of $k = 16$ bits is then $5\times 187\times 16 = 14960$. However, in this experiment, only 10625 bits are needed with our algorithm to reach the desired precision. Our method can still be improved, to use even less information. But note that the additional information represents less than 0.7% (for only 5 partials, though) of the signal information. Such a low information ratio allows the use of watermarking techniques to embed the additional information within the signal itself, as in [1].

[1] Sound examples available on-line at URL: `http://dept-info.labri.fr/~sm/DAFx10/`

## 5. CONCLUSIONS AND FUTURE WORK

In this paper, we have introduced a new method for increasing the precision of the sinusoidal analysis, a way below the Cramér-Rao lower bound (CRB), thanks to some *a priori* knowledge about the sinusoidal parameters to be estimated. This more precise estimation produces sounds of much higher quality, enabling audio effects without typical but unwanted "musical noise".

The present work should be regarded rather as a proof of concept of "informed analysis". Indeed, the informing method is still rather naive, and does not always reach the new informed lower bound (ILB) as it should. Investigating coding and information theories to enhance our method is part of our future research.

However, we have already shown that it is possible to decrease the estimation error of the frequency parameter with at least 2 bits of information, even in the non-stationary case, when the noise level is known though.

Our aim is now to design a complete informed analysis method, including the estimation of the noise level, to fully demonstrate its practical interest. The new method would also be able to inform other sinusoidal parameters, *e.g.* phase and amplitude.

Our long-term perspective is a complete informed analysis / synthesis chain able to create some additional information from the musical content (*e.g.* the several tracks of a musical piece) and use it with the sound signal (*e.g.* the mixed stereo signal of a CD-audio) to allow sound transformations of high quality.

The "informed" concept opens up new research horizons, and should lead for example to new source separation or music information retrieval methods.

## 6. ACKNOWLEDGMENTS

This research was partly supported by the French ANR *(Agence Nationale de la Recherche)* DReaM project (ANR-09-CORD-006).

## 7. REFERENCES

[1] Mathieu Parvaix, Laurent Girin, and Jean-Marc Brossier, "A Watermarking-Based Method for Single-Channel Audio Source Separation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Taipei, Taiwan, April 2009, pp. 101–104.

[2] Robert J. McAulay and Thomas F. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 744–754, 1986.

[3] Xavier Serra and Julius O. Smith, "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic plus Stochastic Decomposition," *Computer Music Journal*, vol. 14, no. 4, pp. 12–24, 1990.

[4] Mathieu Lagrange and Sylvain Marchand, "Assessing the Quality of the Extraction and Tracking of Sinusoidal Components: Towards an Evaluation Methodology," in *Proceedings of the Digital Audio Effects (DAFx) Conference*, Montreal, Quebec, Canada, September 2006, McGill University, pp. 239–245.

[5] Laurent Girin, Sylvain Marchand, Joseph di Martino, Axel Röbel, and Geoffroy Peeters, "Comparing the Order of a Polynomial Phase Model for the Synthesis of Quasi-Harmonic Audio Signals," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, USA, October 2003, pp. 193–196.

[6] Michaël Betser, Patrice Collen, Gaël Richard, and Bertrand David, "Estimation of Frequency for AM/FM Models Using the Phase Vocoder Framework," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 505–517, February 2008.

[7] Sylvain Marchand and Philippe Depalle, "Generalization of the Derivative Analysis Method to Non-Stationary Sinusoidal Modeling," in *Proceedings of the Digital Audio Effects (DAFx) Conference*, Espoo, Finland, September 2008, TKK, Helsinki University of Technology, pp. 281–288.

[8] Kunihiko Kodera, Claude de Villedary, and Roger Gendrin, "A New Method for the Numerical Analysis of Non-Stationary Signals," *Physics of the Earth and Planetary Interiors*, vol. 12, pp. 142–150, 1976.

[9] Kunihiko Kodera, Roger Gendrin, and Claude de Villedary, "Analysis of Time-Varying Signals with Small BT Values," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 64–76, February 1978.

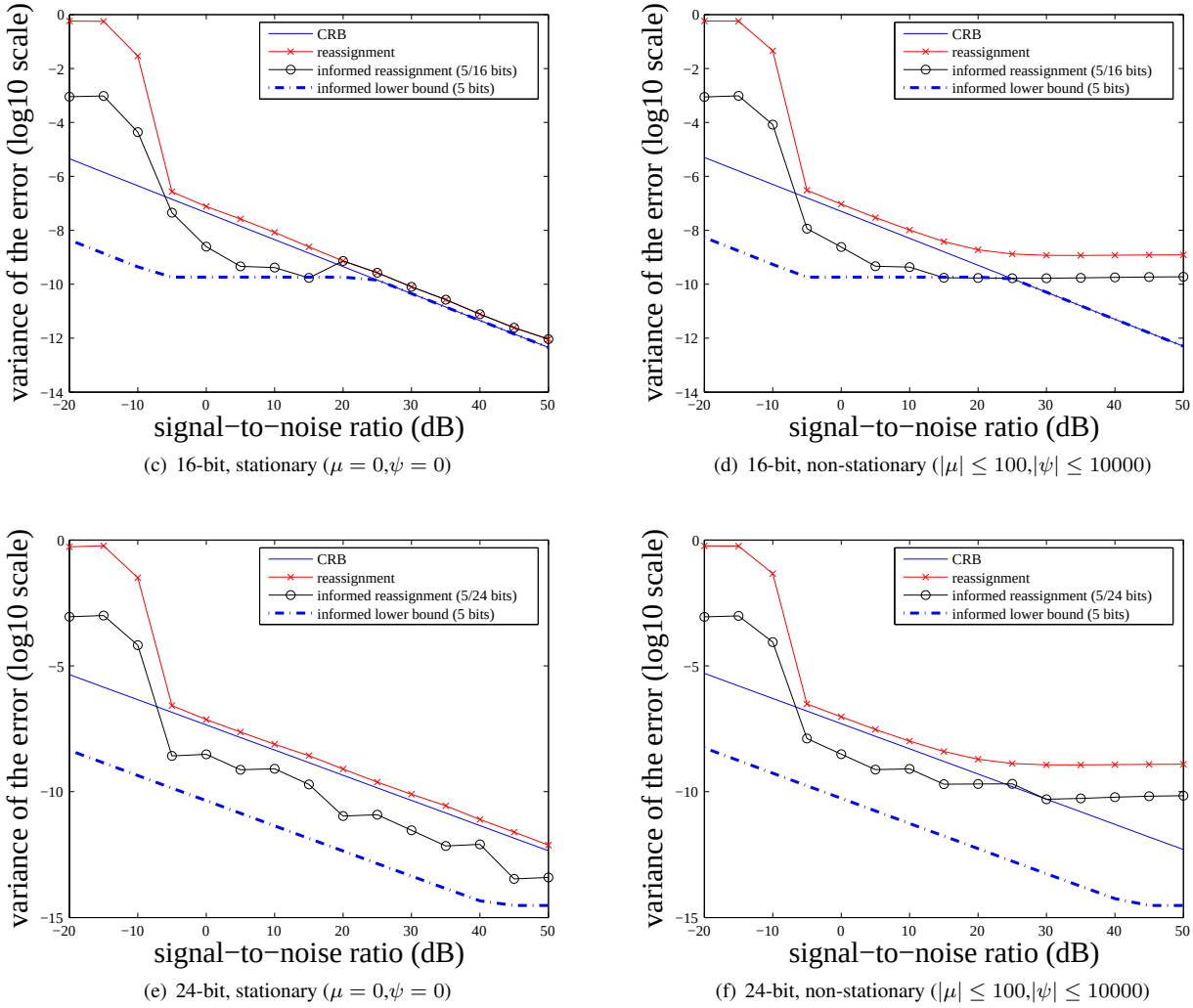
[10] François Auger and Patrick Flandrin, "Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1068–1089, May 1995.

[11] Roland Badeau, Gaël Richard, and Betrand David, "Performance of ESPRIT for Estimating Mixtures of Complex Exponentials Modulated by Polynomials," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 492–504, February 2008.

[12] Guotong Zhou, Georgios B. Giannakis, and Ananthram Swami, "On Polynomial Phase Signal with Time-Varying Amplitudes," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 848–860, April 1996.

[13] Petar M. Djurić and Steven M. Kay, "Parameter Estimation of Chirp Signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 12, pp. 2118–2126, December 1990.

[14] Guillaume Meurisse, Pierre Hanna, and Sylvain Marchand, "A New Analysis Method for Sinusoids+Noise Spectral Models," in *Proceedings of the Digital Audio Effects (DAFx) Conference*, Montreal, Quebec, Canada, September 2006, McGill University, pp. 139–144.



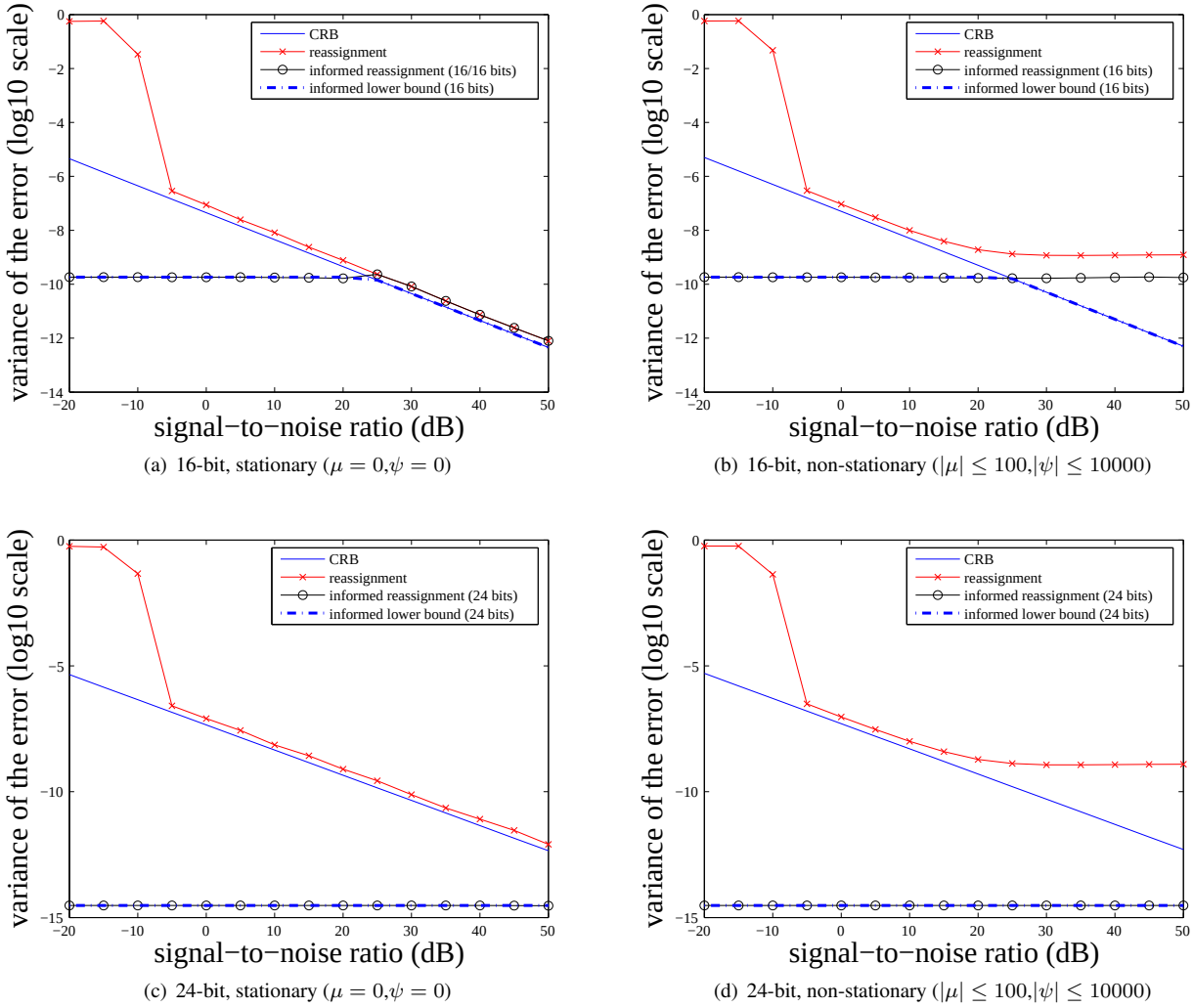

(c) 16-bit, stationary ($\mu = 0, \psi = 0$)

(d) 16-bit, non-stationary ($|\mu| \leq 100, |\psi| \leq 10000$)

(e) 24-bit, stationary ($\mu = 0, \psi = 0$)

(f) 24-bit, non-stationary ($|\mu| \leq 100, |\psi| \leq 10000$)

Figure 5: *Frequency estimation error as a function of the SNR, in several cases (16-bit or 24-bit target precision, stationary or non-stationary cases) for the classic reassignment method and the* ***fixed 5-bit*** *informed reassignment method, and comparison to the Cramér-Rao lower bound (CRB) and the lower bound in the informed case.*




(a) 16-bit, stationary ($\mu = 0, \psi = 0$)

(b) 16-bit, non-stationary ($|\mu| \leq 100, |\psi| \leq 10000$)

(c) 24-bit, stationary ($\mu = 0, \psi = 0$)

(d) 24-bit, non-stationary ($|\mu| \leq 100, |\psi| \leq 10000$)

Figure 6: *Frequency estimation error as a function of the SNR, in several cases (16-bit or 24-bit target precision, stationary or non-stationary cases) for the classic reassignment method and the fully informed reassignment method, and comparison to the Cramér-Rao lower bound (CRB) and the lower bound in the informed case.*